

## Document Metadata Literature Review

ark:/13960/88933fd0-7d40-45fd-8a88-667400a2889e

D:20170530090640+08'00'

What's the cheapest way to achieve the following common document-related tasks: locating, determining the latest revision, and listing the works of an author?

### Digital Object Identifiers

Digital Object Identifiers (DOI) seem popular, and while they are a federated Handle system, the federation charges upkeep fees so that the cheapest document registrar charges 1 USD per document + 275 USD per year.<sup>1</sup>

### Citation Managers

webcite<sup>2</sup> has become for-profit, and neither perma.cc<sup>3</sup> nor archive.is<sup>4</sup> handles PDFs. And none afford the determination of document revision status.

### The Internet Archive

PURL<sup>5</sup> has been adopted by the Internet Archive<sup>6</sup> (IA), however one would have to periodically manage the HTTP 301 redirects as needed.

IA also provides Archival Resource Keys<sup>7</sup> (ARKs) which are similar to DOIs in that one can request the metadata via the ARK Description service, and one can also pull the original resource via the Access service.

Caveats to ARKs include: google does not yet parse ARKs, and it's been an IETF draft since 2001 February.<sup>8</sup>

### urn:uuid:...

As version 4 UUIDs allow for decentralized issuance, generate one UUID to identify the author, and then subsequently issue each paper its own UUID.

Create a PDF/A with the UUID metadata, and submit it to archive.org or webcite. It will then be indexed by google.

Caveats to this include: how does a reader determine which version is latest?

### ARK + UUID

We can self-publish our documents to a high-volume primary distribution facility, and then to IA for permanent storage.

IA allows users to set the ARK identifier on upload<sup>9</sup> which allows the use of locally generated UUIDs as ARKs. The workflow then becomes: create the document with PDF Info metadata and UUID assigned, then upload it to IA. Any time there is a change, publish the new document to IA using the same ARK, and IA will maintain a history of the revisions.

---

<sup>1</sup> <https://www.crossref.org/fees/>

<sup>2</sup> <http://webcitation.org>

<sup>3</sup> <http://perma.cc>

<sup>4</sup> <http://archive.is>

<sup>5</sup> <http://purl.org>

<sup>6</sup> <http://archive.org>

<sup>7</sup> <https://www.ietf.org/archive/id/draft-kunze-ark-18.txt>

<sup>8</sup> <https://tools.ietf.org/html/draft-kunze-ark-00>

<sup>9</sup> <https://blog.archive.org/2013/02/08/presetting-metadata-with-the-new-beta-uploader/>